



La zone préverbale en ancien français : apport des corpus annotés

Thomas Rainsford, Céline Guillot, Alexei Lavrentiev, Sophie Prévost

► To cite this version:

Thomas Rainsford, Céline Guillot, Alexei Lavrentiev, Sophie Prévost. La zone préverbale en ancien français : apport des corpus annotés. 3e Congrès Mondial de Linguistique Française, Jul 2012, Lyon, France. pp.159-176, 10.1051/shsconf/20120100246 . halshs-00758563

HAL Id: halshs-00758563

<https://shs.hal.science/halshs-00758563>

Submitted on 28 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La zone préverbale en ancien français : apport des corpus annotés

T. M. Rainsford

ICAR (CNRS & Université de Lyon)
tmr740-ac@yahoo.co.uk

Céline Guillot

ICAR (CNRS & Université de Lyon)
celine.guillot@ens-lyon.fr

Alexei Lavrentiev

ICAR (CNRS & Université de Lyon)
alexei.lavrentev@ens-lyon.fr

Sophie Prévost

LATTICE (CNRS & ENS)
sophie.prevast@ens.fr

1 Introduction

1.1 L'ancien français : une langue à verbe second (V2)

L'ancien français est généralement considéré comme une langue à verbe second (langue « V2 »), c'est-à-dire dans laquelle le verbe occupe, dans les propositions déclaratives, la seconde position. Il s'agit d'une caractéristique qui rapproche l'ancien français des langues germaniques¹. Ce trait de l'ancien français a été mis en avant depuis longtemps, dans le cadre d'approches très différentes, qu'elles soient simplement descriptives (Le Coultre, 1875 ; Thurneysen, 1892), ou plus théoriques (Foulet, 1930 ; Skårup, 1975 ; Marchello-Nizia, 1995, 2000), ainsi que dans les travaux réalisés dans le cadre de la grammaire générative (par exemple Adams, 1987 ; Vance, 1997).

Si les occurrences de verbe en seconde position sont effectivement majoritaires dans l'ensemble des textes d'ancien français, certains textes offrent néanmoins des exemples de verbes en 1^{ère} position, en particulier les chansons de geste au 12^{ème} siècle :

(1) Plurent Franceis pur pitet de Rollant (*Chanson de Roland*, ca. 1100, v. 3120)

et la plupart des textes présentent des cas de verbes en 3^{ème} position, ou même plus reculés dans la « phrase »². Le dénombrement des éléments qui précèdent le verbe ne tient généralement pas compte de certains éléments : sont exclus ceux qui ne sont pas toniques, tels les pronoms personnels compléments atones (*le, la, les, ...*) ou les conjonctions de coordination, de même que les éléments « hors phrase », telles les apostrophes, les interjections, les incises, ou bien les éléments disloqués. Toutefois, même en écartant ces éléments et en ne considérant que les éléments « toniques » dotés d'un statut argumental, force est de constater que les occurrences de verbes en 3^{ème} position ne sont pas rares, au moins si l'on s'en tient à une perspective descriptive :

(2a) **Li quens Rollant** *Gualter de l'Hum apelet* (*Chanson de Roland*, ca. 1100, v. 803)

- (2b) car **nus** en si haut servise ne doit entrer devant qu'il soit netoiez et espurgiez
de totes vilanies et de toz pechiez mortieus (*Queste del saint Graal*, ca. 1225-
1230, p. 164b³)

Les exemples de ce type conduisent à deux positions quelque peu différentes, quoique conciliables. La première consiste à considérer l'ancien français comme une langue où l'ordre V2 « domine » (voir, entre autres, Marchello-Nizia, 1995). La seconde position consiste, en la résumant brièvement, à postuler une structure profonde de type V2, qui n'est pas nécessairement actanciée en tant que telle en surface. Ainsi, dans le modèle positionnel à trois zones (préverbale, verbale et postverbale) de Skårup (1975), s'ajoute dans certains cas, l'extraposition, à gauche de la zone préverbale, ou une position intercalée (et non comptabilisée) entre les zones préverbale et verbale. L'approche de la grammaire générative, bien que fort différente de la précédente, postule aussi des positions « hors phrase ». Récemment, Rouveret (2004) a cependant proposé une thèse quelque peu divergente, en faisant l'hypothèse, dans le cas d'une périphérie gauche « étoffée », de la coexistence de plusieurs têtes fonctionnelles, qui rendent ainsi disponibles plusieurs positions de spécificateurs.

1.2 Structures V3, V4, ... : une zone préverbale complexe

Sans mettre en cause la pertinence des approches mentionnées ci-dessus, c'est cependant dans une perspective descriptive de surface que nous nous plaçons ici. La question des structures V3, V4, ... (verbe en 3^{ème}, 4^{ème}, ... position), quel que soit le cadre dans lequel elle est abordée, n'a pas donné lieu jusqu'ici à un recensement exhaustif et quantifié, dans un corpus large et diversifié, des éléments qui peuvent précéder le verbe, leur nature et leur fonction. Beaucoup s'en tiennent au constat qu'il existe des dérogations au principe V2, sans approfondir l'importance quantitative du phénomène, ni ses modalités. Cette situation a au moins deux explications. La première réside dans la tendance à s'attacher à la structure dominante — l'ordre V2 — et à délaisser les structures marginales (ou bien à tenter de les ramener à la structure prévalente : démarche de la grammaire positionnelle de Skårup et de la grammaire générative). La seconde raison est d'ordre pratique. En l'absence de corpus enrichi permettant le repérage, l'extraction et le dénombrement automatisés des éléments qui précèdent le verbe, réaliser une telle étude sur un corpus est une entreprise d'une extrême lourdeur. Quelques études se sont néanmoins attachées aux cas de zones préverbaux complexes (« zone préverbale » entendue désormais comme ce qui précède le verbe et les compléments clitiques qui lui sont antéposés, et non dans le sens spécifique que lui donne Skårup). Ainsi en est-il de Marchello-Nizia (2000), qui étudie les éléments préverbaux dans la *Chanson de Roland* et dans la *Queste del saint Graal*⁴. Elle en arrive à la conclusion que, plus que la présence en soi de plusieurs éléments devant le verbe, c'est la cooccurrence de plusieurs éléments thématiques ayant des valeurs fonctionnelles différentes qui fait difficulté (cadre spatial et cadre temporel par exemple). Dans une perspective toute autre, Benincà (2006) propose, dans la lignée de Rizzi (1997), une cartographie détaillée de la périphérie gauche dans les langues romanes médiévales. Elle postule ainsi que l'ordre des éléments initiaux est déterminé par leur fonction pragmatique, les expressions encadrantes précédant les expressions thématiques, qui précèdent les éléments focalisés. Par ailleurs, les « opérateurs sentenciels » (adverbes courts et fréquents : *si, or, ains...*) et les objets doivent apparaître dans le champ « Focus », directement devant le verbe, cette position étant plutôt interne à la « phrase ». Quant aux éléments « encadrants » (entre autres les éléments disloqués), ils sont plutôt externes et ne jouent pas de rôle syntaxique dans la phrase. Sans surprise, Benincà en arrive à la conclusion que l'ancien français évite la multiplication des éléments devant le verbe, contrairement aux dialectes italiens. Mais on peut regretter, pour cette étude, l'absence d'un corpus clairement défini, et celle de quelques données chiffrées.

1.3 Présentation de l'étude

L'annotation morpho-syntaxique et syntaxique réalisée dans le cadre du projet de corpus de référence syntaxique pour le français médiéval (*Syntactic Reference Corpus for Medieval French*, désormais SRCMF)⁵, ainsi que le développement, dans ce même contexte, d'outils d'extraction, a permis dans une large mesure l'automatisation des tâches de repérage, d'extraction et de dénombrement des éléments qui

précèdent le verbe (voir la section 2 pour une présentation détaillée de la méthodologie mise en œuvre dans la présente étude). Il a donc été possible d'envisager, sur un corpus de près de 130 000 mots, une analyse exhaustive et détaillée des éléments qui précèdent le verbe.

L'enjeu de cette étude est de caractériser la structure de la zone préverbale lorsque le verbe n'occupe pas la seconde position, mais une position plus reculée. C'est donc à la configuration syntaxique minoritaire que nous nous attachons, afin de mesurer sa fréquence (est-ce un phénomène réellement marginal ?), et de décrire précisément ses différentes modalités : combien d'éléments peut-on rencontrer devant le verbe ? Quelles sont leurs caractéristiques morphologiques, syntaxiques et sémantiques ? Observe-t-on, au regard de ces différentes dimensions, des associations privilégiées, des schémas récurrents ? Peut-on parler, sinon de règles, au moins de régularités dans les séquences mises au jour ?

Le corpus que nous avons retenu est composé de cinq textes qui présentent des caractéristiques externes différentes, tant en ce qui concerne la date, que la forme (vers/prose) et le domaine textuel. L'étude couvre une période très brève, qui s'étend sur une cinquantaine d'années seulement, de la fin du 12^{ème} siècle au premier tiers du 13^{ème} siècle : la perspective diachronique n'est pas au premier plan. Le corpus comprend deux textes en vers, deux textes en prose, et un texte mixte (alternance de passages en vers et en prose avec une dominance de prose). Trois d'entre eux appartiennent au domaine littéraire (deux romans et un récit bref), tandis qu'un texte relève à la fois du domaine religieux et littéraire (recueil de miracles et de chansons, c'est-à-dire de récits brefs)⁶ et le dernier du domaine historique. Voici, résumées ci-dessous, les caractéristiques des textes du corpus⁷ :

Référence	<i>Beroul</i> ⁸	<i>Coinci</i>	<i>Aucassin</i>	<i>Clari</i>	<i>Graal</i>
Titre et auteur	<i>Tristan de Beroul</i>	<i>Miracles de Gautier de Coinci</i>	<i>Aucassin et Nicolette</i>	<i>La Conquête de Constantinople</i> de R. de Clari	<i>La Queste del saint Graal</i>
Date	entre 1165 et 1200	vers 1218 – 1227	dernier quart du 12 ^e s. ou 1 ^{ère} moitié du 13 ^e s.	après 1205	vers 1225 ou 1230
Forme	vers	vers	mixte	prose	prose
Domaine et genre	littéraire roman	religieux / litt. récits brefs	littéraire récit bref	historique chronique	littéraire roman
Echantillon annoté	intégral	extrait	intégral	intégral	extrait
Nombre de mots	27257	17430	10009	33994	40751
Nombre de phrases	3306	1440	1034	2332	2969

Tableau 1 : Caractéristiques des textes du corpus

Le choix de textes ainsi diversifiés part de l'hypothèse, avérée pour d'autres phénomènes, qu'à des caractéristiques externes différentes peuvent être associées des propriétés langagières différentes. C'est évident en ce qui concerne la date, mais la forme, le domaine et le genre textuels ne sont pas non plus sans incidence. On verra, dans cette étude, que les textes en vers et ceux en prose présentent des structurations de la zone préverbale assez divergentes.

2 Méthodologie

2.1 Modèle linguistique de l'annotation du corpus

Notre analyse se base sur les annotations syntaxiques réalisées dans le cadre du projet SRCMF. Il ne s'agit pas dans cet article de présenter l'ensemble du modèle linguistique et des formalismes adoptés dans ce corpus, mais un bref aperçu des principes méthodologiques ayant un impact direct sur nos résultats, ainsi que des termes utilisés, est indispensable.

Le modèle linguistique adopté dans SRCMF relève de la syntaxe dépendancielle (Tesnière, 1965 ; Polguère & Melčuk, 2009) et tient compte des particularités de la langue médiévale (absence de locuteurs « natifs » et donc impossibilité de recourir à des techniques expérimentales, présence de constructions ambiguës ou ininterprétables, normalisation (discutable) de la ponctuation et de la segmentation des mots dans les éditions modernes de textes médiévaux).

Dans le modèle linguistique SRCMF, l'annotation syntaxique consiste à délimiter des *structures* composées d'une *tête* et de ses *dépendants* éventuels et à étiqueter les relations de dépendance. La tête est toujours un nœud terminal (unité syntaxiquement indivisible), les dépendants peuvent être des structures plus complexes. La structure maximale de l'analyse syntaxique est la *phrase*.

La définition de la phrase est formelle : c'est une structure qui dépend d'un verbe conjugué principal (d'une proposition principale ou d'une « phrase simple »). Chaque verbe principal forme donc une phrase, et la coordination éventuelle à ce niveau de structure syntaxique n'est pas prise en compte. L'exemple (3) donné ci-dessous contient deux phrases selon le modèle linguistique SRCMF et la conjonction *et* est considérée comme un « relateur non coordonnant » :

(3) [Messire Gauvain prend l'espee par le heut _{Phrase}] [et sache _{Phrase}] (*Graal*, p. 161b)

Ce choix méthodologique a pour conséquence une augmentation de la proportion des structures V1 dans le corpus par rapport à ce à quoi conduirait une analyse de la grammaire traditionnelle.

Les énoncés averbaux (très peu nombreux en ancien français) sont exclus de l'analyse. Tous les éléments de la phrase dépendent du verbe principal. Ce sont des *satellites* (sujet, objet, complément, circonstant, etc.) ou des *relateurs* (coordonnants ou non). Chaque satellite est une structure, qui peut être discontinue. Dans le cas d'une proposition subordonnée, il s'agit d'un verbe et de ses propres satellites, dans les autres cas, on identifie simplement la tête et ses *modificateurs* sans préciser davantage le type de la relation qu'ils entretiennent entre eux. Certains satellites sont uniques dans une phrase (sujet, objet, attribut du sujet), d'autres peuvent être multiples (compléments, circonstants). Des éléments ayant une fonction énonciative ou autre sans rapport de dépendance évident avec le verbe (apostrophes, interjections, incidentes) sont néanmoins rattachés au verbe, et regroupés sous le label 'Parenthèse' (Prnt) dans notre analyse. Il convient enfin de noter que la catégorie des circonstants inclut dans le modèle SRCMF des éléments très divers, auxquels nous nous intéresserons de plus près dans la suite de cet article.

Les relateurs (conjonctions ou prépositions) ne sont jamais considérés comme des têtes : par exemple, dans un syntagme prépositionnel, la préposition est un relateur qui dépend du nom qui constitue la tête du syntagme. On ne reconnaît pas d'éléments zéro : des relations n'existent qu'entre des éléments explicites.

La nature des structures syntaxiques (types de syntagmes ou de propositions) n'est pas annotée en tant que telle dans le corpus SRCMF. Néanmoins, grâce à l'étiquetage morpho-syntaxique dont bénéficient les textes du corpus, on peut savoir à quelle partie du discours appartient la tête d'une structure donnée. Il arrive que cette catégorisation sommaire doive être complétée par une analyse formelle et sémantique postérieure plus fine, comme cela a été le cas des circonstants dans le cadre de la présente étude.

Une information absente de l'annotation SRCMF, et pourtant capitale pour une étude qui porte sur l'ordre des mots, est le type d'énoncé (déclaratif, interrogatif ou impératif). Cette annotation supplémentaire a dû être ajoutée manuellement aux résultats des requêtes sur le corpus. En revanche, la structure des vers (et

plus particulièrement la position d'un élément syntaxique par rapport à la fin du vers), qui a une incidence non négligeable sur l'ordre des mots dans les textes poétiques, a pu être récupérée de façon automatique.

Dans le Tableau 2 donné ci-dessous, nous présentons l'ensemble des catégories syntaxiques et morpho-syntaxiques du corpus SRCMF qui apparaissent dans les résultats présentés dans notre article.

<i>Étiquette</i>	<i>Description</i>
ADV	Adverbe
Atbt	Attribut
Circ	Circonstant
Prnt	Parenthèse (interjection, apostrophe, incise ou incidente)
R	Régime du verbe (objet, complément)
Rel	Relateur (conjonction ou préposition)
Sat	Satellite (Sujet, Régime ou Circonstant, etc.)
Sj	Sujet (personnel ou impersonnel)
SN	Syntaxme nominal ou prépositionnel
SUB	Proposition subordonnée
V	Verbe

Tableau 2 : Liste des étiquettes syntaxiques et morpho-syntaxiques citées

2.2 Procédure d'extraction et d'analyse des données

L'interrogation du corpus se fait grâce au logiciel TigerSearch (Lezius, 2002). Cet outil très performant impose toutefois un certain nombre de contraintes, dont il faut tenir compte lors de la préparation du corpus et de la formulation des requêtes. Les structures discontinues et les coordinations demandent une attention particulière.

Les possibilités de visualisation et d'exportation des résultats des requêtes offertes par l'interface graphique de TigerSearch peuvent être augmentées grâce à un post-traitement automatisé à l'aide de scripts. Nous avons ainsi pu construire, grâce à un script Python, des concordances organisées en fonction du nombre d'éléments en position préverbale et des informations pertinentes pour notre analyse.

Compte tenu des contraintes imposées par le modèle linguistique SRCMF et par son outil d'exploitation TigerSearch, la préparation des données analysées dans la présente recherche s'est organisée en deux étapes : 1) extraction des données à partir du corpus avec TigerSearch avec un post-traitement automatisé ; 2) élagage, annotation supplémentaire et analyse à l'aide des logiciels de gestion de tables Microsoft Office Excel et Open Office Calc.

En formulant la requête pour extraire les données pertinentes pour notre étude, nous avons cherché à éviter le « silence » (occurrences pertinentes non repérées) et à réduire le « bruit » (occurrences non pertinentes repérées) dans la mesure du possible. En raison de la complexité du langage de requête, nous ne reproduisons pas ici la requête entière, mais présentons les critères de sélection retenus.

Nous nous intéressons aux phrases qui comportent au moins deux éléments distincts dans la zone préverbale. Certains éléments atones (pronoms personnels régimes, adverbe de négation *ne*, pronoms adverbiaux *en* et *y*) sont généralement considérés comme faisant partie de la zone verbale (Skårup, 1975 ; Marchello-Nizia, 2000). Nous adoptons cette position et ne les prenons pas en compte. Quant aux

éléments « encadrants », que plusieurs chercheurs placent « hors phrase » (Marchello-Nizia, 2000 : 37), nous avons choisi de les comptabiliser tous dans un premier temps, en prenant soin de distinguer trois catégories d'éléments préverbaux :

- Les satellites du verbe (Sat) ;
- Les éléments parenthétiques : apostrophes, interjections, incises et incidentes (Prnt) ;
- Les relateurs de phrase (Rel).

Les résultats de la requête sont exportés dans un format tabulaire où chaque élément en position préverbale et ses annotations (catégorie syntaxique et partie du discours de la tête) sont placés dans des colonnes séparées. Le nombre de colonnes dépend du nombre maximum d'éléments en position préverbale détectés dans chaque texte et ce nombre varie d'un texte à l'autre (de 5 dans *Graal* à 11 dans *Coinci*). Des informations supplémentaires (contextes gauche et droit, références des occurrences, etc.) sont également retenues.

Pour faciliter l'élagage et l'annotation des données, des colonnes de synthèse automatique (par exemple, une reconstitution de chaque phrase avec parenthésage des éléments préverbaux) ont été ajoutées grâce à la fonction 'concaténer' des logiciels tableurs. Des filtres ont été créés afin de pouvoir se concentrer sur des éléments particuliers.

Nous avons ensuite procédé à la vérification et à l'annotation supplémentaire des résultats tirés du corpus. Ont été annotés :

- les types de phrases autres que déclaratives (les énoncés impératifs et interrogatifs ont ensuite été exclus des analyses) ;
- les catégories fines des circonstants (voir la description dans la section 3.2.1).

Une fois la vérification et l'annotation terminées, nous avons généré un certain nombre de colonnes supplémentaires pour faciliter l'analyse des données. Enfin, nous avons utilisé l'outil de rapport de tableau croisé dynamique afin d'effectuer les décomptes présentés plus bas.

3 Analyse des données et résultats

L'analyse des données a été conduite en deux temps. Dans une première phase (section 3.1), nous nous sommes intéressés au nombre d'éléments en position préverbale (désormais EPV) réalisés dans chaque énoncé et à la fréquence des successions d'EPV dans chaque texte. A ce stade, nous avons également comparé les fonctions syntaxiques des combinaisons d'EPV les plus fréquentes dans chaque texte. Dans une seconde phase (section 3.2) nous avons mené une étude beaucoup plus détaillée des circonstants dans le but de mieux décrire leur comportement dans la zone préverbale et leurs possibilités de combinaison dans chaque texte de notre corpus (cf. la présentation des textes du Tableau 1).

3.1 Analyse de toutes les structures d'EPV

L'identification des EPV dans toutes les phrases du corpus nous a permis d'une part de mettre au jour le nombre maximal d'EPV possible dans chaque texte, d'autre part d'étudier la fréquence de chaque structure à n éléments.

Le Tableau 3 présente les résultats de nos dépouillements informatiques texte par texte. Les structures identifiées ici présentent le nombre d'EPV (parenthèses et satellites), en distinguant les cas où un relateur de phrase précède les EPV. En revanche, dans cette estimation globale, ni les types de phrases (déclaratives ou non) ni les types de structures d'EPV (subordonnées, syntagmes, etc.) ne sont spécifiés⁹ :

EPV	Structure	<i>Beroul</i>	<i>Coinci</i>	<i>Aucassin</i>	<i>Clari</i>	<i>Graal</i>
Aucun	V	444	147	114	81	136
	Rel V	68	76	91	378	631
	Total	512 15,5%	223 15,5%	205 19,8%	459 19,7%	767 25,8%
1	EPV V	1902	604	483	692	700
	Rel EPV V	174	140	127	504	823
	Total	2076 62,8%	744 51,7%	610 59,0%	1196 / 51,3%	1523 51,3%
2 ou plus	Total	718 21,7%	473 32,8%	219 21,2%	677 29,0%	679 22,9%
Total		3306	1440	1034	2332	2969

Tableau 3 : Fréquence des différentes structures à *n* EPV dans la phrase

Nous observons d'emblée que la structure à un seul EPV est majoritaire dans tous les textes, ce qui est conforme à nos attentes pour une langue dans laquelle le verbe est censé se trouver en deuxième position dans les déclaratives. Les structures sans EPV sont moins communes que celles à deux EPV ou plus, sauf dans *Graal*. Cependant, dans ce dernier texte, nous observons une proportion très élevée de phrases qui s'ouvrent par un relateur (631 sur 767 occ. de V1). Il s'agit dans l'immense majorité des cas de la conjonction de coordination *et* : il est à rappeler que, dans le corpus SRCMF, chaque verbe principal forme une nouvelle phrase, et la coordination des phrases n'est pas permise (voir l'exemple 3 ci-dessus).

3.2 Analyse des structures à plus de deux EPV

Les énoncés comportant deux EPV ou plus (en comptant les parenthèses et les satellites, mais non pas les relateurs) constituent entre 21,2% et 32,8% des énoncés du corpus. Dans l'analyse qui suit, nous nous focalisons sur ces énoncés qui ont une périphérie gauche plus riche. Le Tableau 4 présente une analyse plus fine de ces structures, grâce à un nettoyage manuel des résultats extraits du corpus. Contrairement au calcul précédent (Tableau 3), seuls les énoncés déclaratifs sont ici pris en compte, et les éléments coordonnés sont traités comme un seul EPV (voir note 9). Les relateurs de phrase ne sont pas comptabilisés comme des EPV.

	<i>Beroul</i>	<i>Coinci</i>	<i>Aucassin</i>	<i>Clari</i>	<i>Graal</i>
Types de structures	81	100	31	33	45
Nombre maximum d'EPV	5	11	6	5	5
Instances de 2 EPV	482 / 78,6%	274 / 66,7%	113 / 77,9%	560 / 87,8%	493 / 78,9%
Instances de 3 EPV	99 / 16,2%	86 / 20,9%	23 / 15,9%	71 / 11,1%	103 / 16,5%
Instances de 4 EPV	27 / 4,4%	34 / 8,3%	7 / 4,8%	6 / 0,9%	27 / 4,3%
Instances de 5 EPV	5 / 0,8%	12 / 2,9%	1 / 0,7%	1 / 0,2%	2 / 0,3%
Instances de 6 et + EPV	0	5 / 1,2%	1 / 0,7%	0	0
Total	613	411	145	638	625

Tableau 4 : Synthèse du nombre de structures ayant au moins deux EPV dans les phrases déclaratives

Ce tableau montre que des différences importantes apparaissent dès lors qu'on s'intéresse de façon fine à toutes les combinaisons d'EPV actualisées dans les différents textes du corpus. Les structures qui sont dénombrées à la première ligne du tableau distinguent la fonction syntaxique des éléments dans leur ordre d'apparition dans la phrase: par exemple la structure 'Sj Circ Circ' à 3 EPV en (4) est distinguée de la structure 'Prnt Prnt Circ' en (5):

- (4) (Qui iroit) (or) (que fous) feroit (*Beroul*, v. 714)
(5) (Dex) (fait li rois) (tant) ai je quis / De lié avoir un sol mesage (*Beroul*, vv. 3404–3405)

Le repérage de toutes ces structures permet d'opposer de manière assez nette les textes en vers et les textes en prose. On observe en effet que le nombre de structures varie de 31 à 45 dans la prose, nombre qui fait plus que doubler (81 à 100) dans les textes en vers. Ces chiffres doivent par ailleurs être pondérés par la taille de chaque texte : les extraits annotés dans *Beroul* et dans *Coinci* ayant un nombre de mots bien moindre que ceux de *Clari* et surtout du *Graal* (plus de deux fois plus de mots que dans *Coinci*), l'écart entre prose et vers n'en est que plus frappant¹⁰. Les résultats présentés dans la section suivante confirmeront à nouveau que les textes en vers comportent des structures beaucoup plus diversifiées que ceux en prose.

Le regroupement des structures par nombre d'éléments situés en position préverbale gomme au contraire ce qui distingue le vers de la prose. En effet, dans tous les textes, les structures les plus fréquentes sont celles qui ont le moins d'EPV devant le verbe : le nombre total de structures à deux EPV avoisine ou dépasse les 70% dans chacune des œuvres (cette fréquence relative est calculée sur le nombre d'énoncés ayant au moins deux EPV et non sur le nombre total d'énoncés du texte). Les énoncés avec plus de trois EPV sont peu nombreux et les structures deviennent relativement atypiques. Enfin, le nombre maximal d'EPV possible est lui aussi relativement constant (entre 5 et 6, à l'exception du texte de *Coinci* : 11).

Une analyse plus fine du détail des structures, en tenant compte de la fonction syntaxique de chaque élément de la zone préverbale, a permis de repérer les combinaisons récurrentes dans chaque texte. Le Tableau 5 donne par ordre de fréquence décroissante la liste des structures qui sont actualisées plus de dix fois dans chacune des œuvres du corpus. Les relateurs ne sont pas comptés comme EPV :

Texte	Structures
<i>Beroul</i>	Circ Circ > Circ Sj > Sj Circ > Sj R > Prnt Sj > Circ R > Prnt Circ > R Circ > R Sj > R R > Circ Circ Circ
<i>Coinci</i>	Circ Circ > Sj Circ > Sj R > Circ R > Prnt Circ > Circ Sj > R Circ > Prnt Sj
<i>Aucassin</i>	Circ Sj > Circ Circ
<i>Clari</i>	Circ Circ > Circ Sj > Circ Circ Circ > Sj Circ > Prnt Sj
<i>Graal</i>	Circ Circ > Circ Sj > Prnt Sj > Prnt Prnt Sj > Sj Circ > Circ Prnt Sj > Prnt Circ > Circ Circ Circ

Tableau 5 : Liste des structures (présentant plus de 10 occurrences) par ordre de fréquence décroissante

Il serait impossible de s'attarder ici sur chacune de ces structures et sur la situation propre à chaque texte. Nous insisterons sur ce qui rapproche et sur ce qui distingue les œuvres, en particulier la prose et le vers. Si les deux structures les plus fréquentes sont à peu près les mêmes dans tous les textes (Circ Circ et Circ Sj), on constate que leur ordre de fréquence relative varie légèrement. Tantôt ce sont les successions de deux circonstants qui arrivent en tête de liste (dans tous les textes sauf *Aucassin*), comme en (6) et (7) ci-dessous, puis les structures contenant un circonstant suivi du sujet (hormis dans *Coinci*), comme en (8) et (9) :

- (6) (Certes) (mot) a li chiens grant sens (*Beroul*, v. 1469)
- (7) (ensi) (si) les prenderoient par forche (*Clari*, p.78)
- (8) (Onques) (li rois) ne s'aperçut (*Beroul*, v. 367)
- (9) (Se jou le coroune) (li empereres Andromes) me tueroit (*Clari*, p. 24)

Tantôt c'est l'ordre inverse (*Aucassin*). Il apparaît par ailleurs qu'en dehors de ces deux structures qui dominent quasiment partout, l'ordre d'apparition des différentes combinaisons d'EPV est relativement variable d'une œuvre à l'autre. De façon générale, on observe que dans tous les textes le sujet et les circonstants sont les éléments les plus présents dans la zone préverbale, les autres actants du verbe (regroupés dans la catégorie régime) étant plus rarement placés dans cette position. On peut aussi souligner le fait que, dans la quasi-totalité des œuvres, les structures à deux EPV les plus fréquentes ne contiennent pas le sujet mais uniquement des circonstants ('Circ Circ'). On note également que, dans les deux textes en vers du corpus, la structure 'Sj Circ' fait partie des structures les plus fréquentes, comme par exemple en (10) :

- (10) (L'ame) (sovent) pleure (*Coinci*, I Pr 2, v. 67)

Dans *Coinci*, elle est même davantage utilisée que la structure 'Circ Sj', qui est au contraire très représentée dans la prose (on reviendra sur ce point dans la section suivante). On constate enfin que toutes les parenthèses sont le plus souvent en début d'énoncé et s'intercalent assez peu entre deux éléments appartenant à la structure argumentale du verbe.

Si la nature des éléments et les types de combinaisons les plus fréquents varient somme toute assez peu d'un texte à l'autre, la fréquence de ces éléments et des combinaisons se révèle sensiblement moins stable. Le tableau 6, ordonné par type de structure et non plus par fréquence décroissante, permet de comparer plus facilement les fréquences texte par texte. Les pourcentages sont établis à partir du nombre total d'énoncés ayant au moins deux EPV dans chaque texte :

	<i>Beroul</i>	<i>Coinci</i>	<i>Aucassin</i>	<i>Clari</i>	<i>Graal</i>
Circ Circ	99 / 16,4%	50 / 12,4%	37 / 26,4%	421 / 66,3%	258 / 41,5%
Circ Circ Circ	11 / 1,8%	2 / 0,5%	0	40 / 6,3%	11 / 1,8%
Circ Circ Sj	4 / 0,7%	4 / 1%	4 / 2,9%	5 / 0,8%	10 / 1,6%
Circ Prnt Sj	8 / 1,3%	1 / 0,2%	2 / 1,4%	3 / 0,5%	15 / 2,4%
Circ R	40 / 6,6%	30 / 7,5%	4 / 2,9%	2 / 0,3%	10 / 1,6%
Circ Sj	76 / 12,6%	17 / 4,2%	43 / 30,7%	63 / 9,9%	134 / 21,5%
Prnt Circ	35 / 5,8%	19 / 4,7%	3 / 2,1%	6 / 0,9%	15 / 2,4%
Prnt Prnt Sj	5 / 0,8%	2 / 0,5%	10 / 7,1%	8 / 1,3%	29 / 4,7%
Prnt R	15 / 2,5%	6 / 1,5%	2 / 1,4%	1 / 0,2%	7 / 1,1%
Prnt Sj	40 / 6,6%	12 / 3%	6 / 4,3%	16 / 2,5%	31 / 5,0%
R Circ	20 / 3,3%	16 / 4%	2 / 1,4%	9 / 1,4%	1 / 0,2%
R R	13 / 2,2%	5 / 1,2%	2 / 1,4%	0	0
R Sj	14 / 2,3%	7 / 1,7%	1 / 0,7%	2 / 0,3%	2 / 0,3%
Sj Circ	48 / 8%	47 / 11,7%	3 / 2,1%	29 / 4,6%	17 / 2,7%
Sj Circ Circ	4 / 0,7%	10 / 2,5%	0	3 / 0,5%	2 / 0,3%
Sj Circ R	5 / 0,8%	10 / 2,5%	0	0	0
Sj R	46 / 7,6%	35 / 8,7%	2 / 1,4%	1 / 0,2%	1 / 0,2%
<i>Autres</i>	<i>119 / 19,8%</i>	<i>129 / 32,1%</i>	<i>19 / 13,57%</i>	<i>26 / 4,1%</i>	<i>81 / 13,0%</i>
Total	602 / 100%	402 / 100%	140 / 100%	635 / 100%	625 / 100%

Tableau 6 : Liste et fréquence des structures alignées texte par texte

Ce tableau fait apparaître des disparités très importantes d'un texte à l'autre. Si l'on s'en tient à la structure la plus représentée dans chaque texte, il apparaît qu'elle correspond à 66,3% des énoncés dans *Clari*, 41,5% dans le *Graal*, 30,7% dans *Aucassin*, mais seulement à 12,4% des énoncés dans *Coinci* et à 16,4% dans *Beroul*. Ces chiffres correspondent à la structure 'Circ Circ' dans tous les textes hormis *Aucassin*, où l'ordre Circ Sj est dominant. Si l'on observe la fréquence de cette dernière structure, les fréquences sont également très différentes d'un texte à l'autre : de 30,7% dans *Aucassin* à 4,2% dans *Coinci*, en passant par 21,5% dans *Graal*, 12,6% dans *Beroul* et 9,9% dans *Clari*. On observe par ailleurs que la catégorie « autres », qui regroupe toutes les structures rares, correspond dans les deux textes en vers (et uniquement dans ceux-ci) à la part la plus importante des énoncés. La grande variété des structures représentées dans le vers est bien entendu à mettre en relation avec le fait qu'aucune d'entre elles ne l'emporte massivement sur les autres.

Après avoir établi ces quelques données générales, nous avons choisi d'examiner plus en détail les circonstants, massivement représentés dans les EPV et qui se combinent tantôt avec les arguments du verbe, tantôt entre eux. Ce sont donc les différentes possibilités d'association des circonstants que nous étudierons dans la section suivante.

3.3 Analyse des circonstants en position préverbale

3.3.1 Classement des circonstants

Dans le but d'observer plus précisément les circonstants dans toutes les constructions qui contiennent au moins deux éléments préverbaux, nous avons procédé à un classement plus fin des formes de ces circonstants en distinguant en premier lieu trois grandes catégories : 1) les adverbes, 2) les syntagmes nominaux au sens large (avec ou sans préposition), 3) les propositions subordonnées, participiales ou infinitives.

A l'intérieur de la catégorie des adverbes, nous avons regroupé les adverbes « qualificatifs » (en *-ment*), les adverbes locatifs (*ci*, *là*, *dessus*, etc.) et les adverbes temporels. Pour les autres adverbes, plus difficiles à classer, nous nous sommes limités à une lemmatisation permettant de neutraliser les variations graphiques (par exemple l'adverbe SI peut être graphié *si*, *se* ou *s'* selon les contextes dans certains textes).

Parmi les syntagmes nominaux, nous avons distingué les locutions ayant une valeur temporelle ou spatiale et les syntagmes à fonction énonciative (*Par Dieu*, etc.). Le reste a été regroupé dans une grande catégorie.

Pour les subordonnées, nous avons distingué les temporelles en *quant*, les autres temporelles, les hypothétiques en *se* (conjonction 'si' en français moderne). Les autres subordonnées ont été regroupées. Les propositions infinitives et participiales sont, bien entendu, classées à part.

Comme nous l'avons déjà constaté lors de l'analyse globale des fréquences, une différence entre la prose et le vers apparaît de manière manifeste. De façon générale, les types de constructions dominants se dessinent plus nettement en prose, tandis que la diversité est plus importante dans les textes en vers. Pour cette raison, nous commencerons la présentation des résultats par les tendances qui se dégagent de l'analyse des textes en prose, avant de procéder à une comparaison avec les textes en vers.

3.3.2 Combinaisons de plusieurs circonstants en position préverbale

Parmi les combinaisons de deux EPV ou plus devant le verbe, c'est la structure qui comporte deux circonstants en position préverbale qui est la plus fréquente dans tous les textes, sauf dans *Aucassin* où cette structure arrive en deuxième position.

C'est dans *Clari* que la dominance de cette structure est la plus marquée avec 66% des phrases ayant au moins deux EPV. Un examen plus détaillé montre que dans la grande majorité des cas le deuxième circonstant (le plus proche du verbe) est l'adverbe *si* (411 occ. sur 425). Au sein de cet ensemble, une structure particulière est utilisée d'une façon extrêmement fréquente : une subordonnée temporelle en *quant* (278 occ.) précède l'adverbe *si*, comme dans l'exemple suivant :

- (11) Et (quant il furent revenu) (si) fissent savoir as barons qu'il avoient fait (*Clari*, p. 5)

Les autres types de subordonnées apparaissent dans 30 cas seulement. L'omniprésence dans ce texte de la structure *quant* + *si* avait été observée déjà par Schon (1960).

Dans 85 énoncés, ce n'est pas une proposition mais un adverbe qui apparaît devant *si*. Il s'agit le plus souvent de *adont* (31 occ.), *après* (21 occ.) ou *puis* (18 occ.) :

- (12) (Adont) (si) manda li dux tous les haus consaus de le vile (*Clari*, p. 7).
(13) (Après) (si) se fist coroner tot a force a empereur (*Clari*, p. 20)
(14) et (puis) (si) les fist geter en le mer (*Clari*, p. 20)

Dans ce texte, les structures qui comportent trois circonstants préverbaux ne sont pas exceptionnelles (37 occ.). Il s'agit généralement de la structure dominante de ce texte (subordonnée en *quant* + *si*) renforcée

par un adverbe comme *adont*, *donc*, *puis* ou *après* (24 occ.). Ces adverbes peuvent être placés avant ou après la subordonnée, l'adverbe *si* précède toujours immédiatement le verbe :

- (15) (Quant li dux de Venice eut si faitement parlé) (après) (si) dist as barons qu'il esleussent les leur dis que il aroit molt tost eslit les siens dis (*Clari*, p. 91)
- (16) (Après) (quant tout chou fu fait) (si) fu li Noeux passés (*Clari*, p. 69)

Dans un cas, nous avons une structure dans laquelle on peut identifier quatre circonstants :

- (17) (Quant uns enfers hons qui avoit mal dedens le cors si comme d'enfle qui dedens le ventre estoit enflés le metoit en se bouche) (ja si peu ne li eust mis) (quant chus buhotiaus le prenoit) (se) li suchoit toute chele maladie (*Clari*, p. 85)

Il s'agit en réalité d'une longue subordonnée temporelle (avec deux relatives enchâssées) suivie d'une concessive, suivie d'une nouvelle subordonnée temporelle qui résume la première de manière anaphorique, suivie enfin de l'adverbe *si*, quasi-obligatoire dans ce texte pour introduire une principale après une subordonnée antéposée.

Il convient de noter qu'aucune subordonnée n'est placée directement devant le verbe. Il faut tout de même relativiser cette affirmation, dans la mesure où nous n'avons pas analysé les cas où le circonstant est le seul élément préverbal. En revanche, nous avons pu vérifier qu'une subordonnée peut être placée devant le verbe lorsqu'un relateur se trouve en début de phrase. Ce type de construction est rare, mais pas inexistant (3 occ.). Une analyse plus fine montre tout de même que des éléments « non comptabilisés » se glissent dans les trois cas entre la subordonnée et le verbe : un adverbe de négation, le pronom adverbial *en* ou *y* :

- (18) et (quant il vint de Jherusalem en Coustantinoble) **n'en** avoit il que deus vis (*Clari*, p. 55)

Dans *Graal*, la dominance de la structure 'Circ Circ' est moins marquée que dans *Clari*, mais elle est tout de même incontestable avec 41,5% des phrases ayant au moins 2 EPV (258 occ.). Comme dans *Clari*, pour la grande majorité de ces occurrences (85%), il s'agit de l'adverbe *si* situé directement devant le verbe et précédé d'un autre circonstant. A quelques exceptions près, la construction subordonnée temporelle en *quant* + *si* est la seule attestée (213 occ. sur 220). A la différence de *Clari*, les constructions où un autre adverbe (cf. 12–14) ou un syntagme nominal se place devant *si* sont extrêmement rares (2 et 3 occ. respectivement).

Les constructions où l'on retrouve trois circonstants en position préverbale sont également nettement plus rares dans *Graal* (11 occ.) que dans *Clari*. Pour 9 d'entre elles, il s'agit d'un élément temporel (adverbe, syntagme nominal ou subordonnée) suivi d'une subordonnée temporelle en *quant* suivie à son tour de l'adverbe *si*, comme dans les exemples suivants :

- (19) et (puis) (quant il est vestuz et apareilliez) (si) le viennent veoir tuit cil de laienz (*Graal*, p. 221c)
- (20) (Au matin) (quant il s'esveilla) (si) resgarda tot entor soi (*Graal*, p. 218d)

Aucune subordonnée circonstancielle n'est placée directement devant le verbe dans les phrases déclaratives de ce texte.

Dans *Aucassin*, la structure 'Circ Circ' représente 26,4% des énoncés ayant au moins deux EPV. Parmi ces 37 phrases, la structure de loin la plus fréquente comprend une subordonnée suivie d'un adverbe (*si*, *or*, *ja*, *donc*, *encore*, *jamais*) (25 phrases).

- (21) (quant ele voit son ami) / (or) fu lie (*Aucassin* [vers], p. 39)

Il n'y a aucun exemple dans ce texte où la subordonnée est le deuxième EPV directement devant le verbe. Les autres exemples de la combinaison de deux circonstants comprennent une séquence de deux adverbes

(5 cas, dont 3 de *puis si*), un SN et un adverbe (4 cas, dont deux avec un SN à valeur énonciative du type *par Diu de maïesté*), deux SN (3 cas) et un adverbe et un SN (2 cas).

La succession de deux circonstants constitue dans le vers, comme dans la plupart des autres textes, la structure à deux EPV la plus répandue. On observe également qu'elle est le plus souvent réalisée par la succession 'SUB ADV' (26 occ¹¹ dans *Beroul*, 14 dans *Coinci*). Mais ce qui distingue nettement les textes en vers, c'est d'une part que la fréquence de la structure à deux circonstants reste très modérée (16,4% et 12,4% sur le total des énoncés à deux EPV et plus), d'autre part le fait qu'elle s'actancie dans un ensemble de combinaisons beaucoup plus diverses que ce n'est le cas dans la prose :

(22) (Qant en voi un qui el tai voire) / (Adonc) flavele cil a cuite (*Beroul*, vv. 3685–6)

(23) (s'il or m'en ot parler) / (Bien) me porroit tenir por fole (*Beroul*, vv. 176–7)

Les adverbes qui suivent la subordonnée sont très variés — voir, par exemple, *adonc* dans l'exemple (22) et *bien* dans l'exemple (23) — et l'adverbe *si* n'est utilisé qu'une fois dans *Beroul* et jamais dans *Coinci*. Les subordonnées sont également variées, et les temporelles ne dominent pas comme elles le font dans *Clari* et *Graal* (cf. la subordonnée conditionnelle en *se* dans l'exemple (23)). Malgré toutes ces variations, on observe qu'il est relativement rare qu'une subordonnée précède directement le verbe et c'est le plus souvent le sujet qui s'intercale entre les deux éléments :

(24) (Se de la mere Dieu chantez) / (Tous enchantanz) iert enchantez (*Coinci*, I Ch 3, v. 10)

Dans quelques cas cependant, dans *Beroul* (10 occ) plus que dans *Coinci* (7 occ), la structure 'SUB (généralement concessive) V' se rencontre :

(25) (Rois) (s'or n'en prens aspre venjance) / n'as droit en terre sanz doutance (*Beroul*, vv. 1903–4)

On constate également que, dans les deux œuvres en vers du corpus, la structure 'ADV ADV' n'est pas beaucoup moins fréquente que la structure 'SUB ADV'. Elle est réalisée dans 16 énoncés dans *Beroul* et dans 19 énoncés dans *Coinci*. Les combinaisons d'adverbes du type *ainc puis*, *puis si*, *ainsi si* se rencontrent de manière sporadique, comme par exemple en (26) :

(26) N'(ainc) (puis) n'i peut goute veoir (*Coinci*, I Pr 1, v. 242)

D'autre part, les adverbes énonciatifs comme *certes*, *nequedent*, *neporquant*, *ja* sont généralement les plus éloignés du verbe :

(27) Et (nequeden) (si) ot s'espee / Entor ses flans estroit noee (*Beroul*, v. 3575–6)

3.3.3 Circonstants et autres arguments

La combinaison 'Circ Sj' s'avère être la plus fréquente dans *Aucassin*, avec 43 énoncés (30,7% des phrases ayant au moins deux EPV). Dans la majorité de ces énoncés, c'est une subordonnée qui précède le sujet (32 cas) :

(28) et (s'il vient ci) (nos) li dirons (*Aucassin* [prose], p. 20)

La structure 'Adverbe Sj' est cependant bien attestée (11 cas). Les adverbes utilisés ont plutôt une fonction connectrice (*ainc*, *donc*, *neporquant*) ou énonciative (*certes*). Il y a par ailleurs quatre occurrences de *ja* dans les expressions *ja Dix ne m'aït* et *ja Dix ne me doinst riens*.

Cette même combinaison 'Circ Sj' est presque aussi fréquente dans *Graal*, où elle représente 21,8% des cas (134 occurrences). Comme dans *Aucassin*, c'est le plus souvent une subordonnée qui précède le sujet (101 des 134 cas), avec une prévalence des subordonnées temporelles (67 occurrences ; voir l'exemple (29)) et dans une moindre mesure des hypothétiques (29 cas ; voir l'exemple (30)) :

(29) Et (quant ele ot grant piece langui) (nos) resgardasmes quel maladie ele avoit
(*Graal*, p. 216d)

(30) car (se tu te prenoies a lui par bataille) (tu) sez bien (*Graal*, p. 192d)

La structure 'ADV Sj' n'est pas rarissime : elle se rencontre dans 22 cas (16,4% des 134 cas) ; sans surprise on ne rencontre pas l'adverbe *si* conjointement avec un sujet préverbal.

La combinaison 'Circ Sj' est bien moins fréquente dans *Clari*, elle ne représente que 9,9% des cas (63 occurrences). Dans près de la moitié des occurrences (30), le sujet est précédé d'un adverbe (dont *après* à 14 reprises) :

(31) (Aprés) (li message) prisent congié (*Clari*, p. 8)

Le sujet est souvent précédé aussi d'une subordonnée (29 cas), temporelle ou hypothétique (dans une proportion voisine) :

(32) (s'il kemandaissent « Alés le pas ») (on) aloit le pas (*Clari*, p. 47)

Il faut noter que, à partir de cette combinaison de base, on peut rencontrer des combinaisons plus complexes, avec plusieurs Circ devant le sujet et/ou l'insertion de parenthèses. Ainsi, dans *Clari*, aux 63 occurrences de 'Circ Sj', s'ajoutent 12 occurrences d'une structure plus complexe, qui se décline selon 4 schémas, par exemple 'Circ Prnt Prnt Sj'¹² :

(33) (Chertes) (seigneur) (fist-il) (vous) avés fait molt grant barnage et molt grant
proeche ... (*Clari*, p. 53)

Dans *Graal*, aux 134 occurrences de 'Circ Sj', il faut ajouter 63 occurrences d'une structure plus complexe, qui se décline sous la forme de 11 schémas différents, avec une prévalence de : 'Circ Prnt Sj' (15 occurrences) :

(34) (Par foi) (fait ele) (je) le vos dirai voiant toz cels de ceste place (*Graal*, p.
162d)

Dans *Aucassin*, à part les combinaisons 'Circ Circ' et 'Circ Sj', il n'y a que 17 exemples de structures plus complexes qui ont au moins un Circ séparé du verbe par un autre EPV¹³. Comme dans les structures précédentes, le Circ est souvent une subordonnée (12 exemples) ou un adverbe (5 cas, dont 4 de *certes*). Plus rarement, un SN à valeur énonciative peut être séparé du verbe, y compris par une parenthèse :

(35) (Par foi) (fait il) (les deniers) prenderons nos (*Aucassin* [prose], p. 20)

Les exemples d'un Circ (ou plus) qui se place entre un actant et le verbe sont globalement assez rares. Dans *Aucassin*, il n'y en a que 5 dans tout le texte (10,6%). A trois reprises, un seul Circ s'intercale entre le sujet et le verbe, dont une occurrence de la structure figée du type *il si fist*. On a par ailleurs deux cas de 'R Circ', l'élément régime étant un verbe à l'infinitif, et la structure se trouvant dans la partie en vers du texte :

(36) (a demanter) (si) se prist / si con vos porrés oïr (*Aucassin* [vers], p. 12)

Le Circ qui s'intercale est toujours un adverbe court (*or*, *si*).

Dans *Clari*, les occurrences d'un Circ placé entre un actant et le verbe sont moins fréquentes, puisqu'elles représentent 6,5 % (41 cas). Le plus souvent (38 cas), un seul Circ est intercalé, parfois entre le régime et le verbe (9 occurrences), le plus souvent (29 cas) entre le sujet (nominal ou pronominal) et le verbe. Dans ce dernier cas, le circonstant est l'adverbe *si* dans 27 cas :

(37) (Ches deux ymages) (si) seoient devant le cange qui molt soloit estre rikes
illuec (*Clari*, p. 87)

(38) (et) (il) (si) fisent (*Clari*, p. 25)

Il convient de noter que la construction figée *il si fisent* représente 9 des 10 occurrences où le sujet séparé du verbe est pronominal. Dans les 3 occurrences où 2 Circ sont intercalés entre le Sujet et le verbe, c'est selon un même schéma : 'Sujet + Subordonnée temporelle en *quant* + adverbe *si*'.

Dans *Graal*, la fréquence des occurrences d'un Circ intercalé entre un actant et le verbe est particulièrement faible : 2,9% (18 occurrences). Le plus souvent un seul Circ sépare l'actant de son verbe, et celui-ci est majoritairement un sujet. L'adverbe *si* est bien représenté, que le sujet soit nominal ou pronominal.

C'est peut-être dans ce type de construction, où le circonstant se combine à un argument du verbe, que le vers se distingue le plus nettement de la prose. Certes, la structure 'Circ Sj' est bien représentée ici aussi (18 occ. dans *Coinci* et 28 si on comptabilise les structures à plus de deux EPV ; 84 occ dans *Beroul* et plus de 100 si l'on comptabilise les structures à plus de deux EPV), et ce sont généralement les subordonnées qui sont suivies du sujet puis du verbe, comme dans les exemples en prose (28–30). Mais il arrive aussi assez souvent que ce soit un élément régime qui se place entre le circonstant et le verbe (30 occ. dans *Coinci* et 40 dans *Beroul*), et il arrive même que le sujet se trouve devant cet ensemble en début d'énoncé comme dans la structure 'Sj Circ R' (10 occ. dans *Coinci*) :

- (39) (Orgiuelz) (tous jors) (tot son sanc) sue / Quant voit nului qui le sourmonte
 (*Coinci*, I Mir 10 : 1946-7)

L'autre caractéristique du vers est qu'il permet beaucoup plus facilement que la prose l'antéposition du sujet au circonstant. Dans *Beroul*, 48 énoncés sont de ce type et dans *Coinci* 47. Dans *Beroul* surtout, on constate que le circonstant est alors majoritairement un SN (29 des 48 occ.) et plus rarement un adverbe :

- (40) (Tristran) (du cri qu'il ot) s'esvelle (*Beroul*, v. 2077)

Enfin, la structure 'R Circ', dans laquelle un Circ sépare l'objet ou un complément du verbe principal, est loin d'être absente dans les deux textes en vers, puisqu'elle est réalisée dans 20 énoncés dans *Beroul* et dans 16 énoncés dans *Coinci* (et 23 si l'on y ajoute les structures à plus de deux EPV). On observe qu'il est assez fréquent que le verbe soit alors en fin de vers. C'est le cas par exemple dans toutes les structures 'R Circ' avec adverbe temporel dans *Beroul* :

- (41) (La meslee) (des or) vos vié (*Beroul*, v. 3130)

Il semble ainsi que la position du verbe (en début et fin de vers, ou en position médiane), puisse influencer sur le reste de la structure du vers. En particulier, quelques sondages ont montré que bon nombre des structures qui se rencontrent dans le vers, et peu ou pas dans la prose, partagent cette particularité que le verbe s'y trouve en fin de vers. Dans quelques cas, le verbe se trouve également en fin de phrase, ce qui a pour conséquence que la périphérie gauche correspond à toute la phrase :

- (42) (Et) (quant je Dé proié avrai) / (A vos) (eisinc) (lors) revendrai (*Beroul*, vv. 937–8)

On voit ici l'influence que les contraintes métriques peuvent avoir sur l'ordre des mots dans le vers. Il serait assurément très utile de prendre en compte ces contraintes de manière plus systématique, et c'est ce qu'on se propose de faire dans une prochaine suite de cette étude.

4 Conclusion

L'approche plutôt descriptive que nous avons adoptée dans cet article nous a permis de mettre au jour des régularités dans l'ordre des constituants dans la périphérie gauche en ancien français, tout en soulignant des différences importantes, dans notre corpus, entre le vers et la prose.

Dans un premier temps, nous avons vu que les structures dans lesquelles plusieurs éléments précèdent le verbe fini sont présentes dans tous les textes, bien que la structure à un seul EPV, ordre V2 canonique, soit toujours majoritaire (Tableau 3). Dans les textes en prose de notre corpus (*Clari*, *Graal* et, pour la plus grande partie, *Aucassin*), la majorité des structures à deux EPV et plus correspond à un ensemble assez limité d'exceptions déjà repérées à la règle V2, notamment, la présence d'un circonstant (souvent

si) ou du sujet entre une subordonnée et le verbe, ou la tendance à situer les circonstants à valeur énonciative (*par foi*, etc.) et les adverbes connecteurs (*ainc*, *puis*) au début de la périphérie gauche. Il est à noter que la langue du seul texte non-littéraire du corpus (*Clari*) est aussi la plus stéréotypée, avec une fréquence très élevée de la structure subordonnée temporelle + *si*. Si l'on accepte de considérer l'élément initial dans ces structures comme étant extraposé, ces textes soutiennent bien l'hypothèse que l'ancien français est une langue à V2 assez régulière.

Dans les textes en vers, même si les structures avec deux EPV ou plus ne sont pas plus fréquentes dans leur totalité que dans les textes en prose, la variété des structures attestées est beaucoup plus importante. On observe néanmoins, dans une certaine mesure, les mêmes phénomènes que dans les textes en prose : par exemple, l'ordre subordonnée + X est beaucoup plus commun que l'ordre X + subordonnée ; certains circonstants adverbiaux (*si*, *bien*, *moult*, *mieux*) sont toujours à côté du verbe ; les syntagmes énonciatifs se placent normalement le plus à gauche. Mais on rencontre beaucoup plus souvent l'antéposition de deux satellites, notamment les structures 'Sj Circ', 'Sj R', et 'Circ R'. Il est particulièrement intéressant d'observer la fréquence de l'antéposition d'un régime dans les textes en vers, même au XIII^e siècle (*Coinci*), dans la mesure où Marchello-Nizia (1995) observe que l'antéposition de l'objet est nettement moins fréquente durant cette période. Cependant, son étude se base sur une comparaison du vers du début du XII^e (*La Chanson de Roland*) avec la prose du XIII^e siècle (*Graal*). S'il était avéré que les contraintes métriques (et particulièrement la position du verbe en fin de vers) peuvent parfois favoriser ce type d'antéposition, il s'ensuivrait que la comparaison du vers et de la prose devrait prendre en compte ces facteurs métriques. Une étude plus précise de l'influence des facteurs métriques sur l'ordre des éléments reste donc à mener dans de futures recherches.

Références bibliographiques des textes étudiés

- Bérout¹⁴, *Tristan*, éd. E. Muret et L. M. Defourques, Paris : Champion, 1947.
- Gautier de Coinci, *Miracles*, t. 1, éd. V. F. Koenig, Genève : Droz, 1955.
- Aucassin et Nicolette, éd. M. Roques, Paris : Champion, 1929.
- Robert de Clari, *La Conquête de Constantinople*, éd. P. Lauer, Paris : Champion, 1924.
- La *Queste del saint Graal*, éd. Ch. Marchello-Nizia, Lyon : ENS de Lyon, 2011, prototype d'édition numérique <http://txm.risc.cnrs.fr/txm>.

Références linguistiques

- Adams, M. (1987). *Old French, Null Subjects and Verb Second Phenomena*, Ph.D. Dissertation, University of California, Los Angeles.
- Benincà, P. (2006). A detailed map of the left periphery of Medieval Romance. In Zanuttini, R., Campos, H., Herburger, E. et Portner, P. H. (éd.) *Georgetown Round Table on Languages and Linguistics (2004). Crosslinguistic Research in Syntax and Semantics : Negation, Tense and Clausal Architecture*. Washington, D.C. : Georgetown University Press, 53-86.
- Buridant, C. (1987). Les résidus de l'ordre OV en ancien français et leur effacement en moyen français. *Romania*, 108, 20-65.
- Combettes, B. (1985). *Recherches sur l'ordre des éléments de la phrase en moyen français*, Thèse pour le Doctorat d'Etat, Université de Nancy, exemplaire dactylographié.
- Donaldson, B. (à par.). Les subordonnées initiales et la périphérie gauche en ancien français. In Combettes, B., Guillot, C., Prévost, S., Oppermann-Marsaux, E., Lavrentiev, A. (éd.), *Le changement en français. Etudes de linguistique diachronique, V*, Bern : Peter Lang.
- Foulet, L. (1930). *Petite syntaxe de l'ancien français*. Paris : Champion.

- Lavrentiev, A. (2010). La 'phrase' en français médiéval : une réalité ou une reconstruction artificielle ? In Neveu, F., Muni Toke, V., Durand, J., Klinger, T., Mondada, L., Prévost, S. (éd.), *Actes du 2^e Congrès Mondial de Linguistique Française*, La Nouvelle Orléans, 12-15 juillet 2010, 277-289.
- Le Coultre, J. (1875). *De l'ordre des mots dans Crestien de Troyes*. Leipzig : Teubner.
- Lezius, W. (2002). Ein Suchwerkzeug für syntaktisch annotierte Textkorpora (German). *University of Stuttgart Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS)*, vol. 8, no. 4. Institut für Maschinelle Sprachverarbeitung (IMS), Stuttgart.
- Marchello-Nizia, C. (1995). *L'évolution du français : ordre des mots, démonstratifs, accent tonique*. Paris : Armand Colin.
- Marchello-Nizia, C. (2000). Le décumul du thème dans l'évolution du français, *Le Français Moderne*, 1, 31-40.
- Polguère, A., Mel'čuk, I. (éd.) (2009). *Dependency in linguistic description*. Amsterdam, Philadelphia : Benjamins.
- Rizzi, L. (1997). The fine structure of the left periphery. In Haegeman L. (éd.) *Elements of grammar, Handbook of generative syntax*, Dordrecht : Kluwer, 281-337.
- Rouweret, A. (2004). Les clitiques pronominaux et la périphérie gauche en ancien français. *Bulletin de la Société de Linguistique de Paris*, t. XCIX, fasc. 1, 181-237.
- Schon, P. M. (1960). *Studien zum Stil der frühen französischen Prosa. Robert de Clari, Geoffroy de Villehardouin, Henri de Valenciennes*, Frankfurt-am-Main : Vittorio Klostermann.
- Skårup, P. (1975). *Les premières zones de la proposition en ancien français. Essai de syntaxe de position*, København : Akademisk forlag, *Revue Romane*, numéro spécial, 6.
- Tesnière, L. (1965). *Éléments de syntaxe structurale*, Paris : Klincksieck.
- Thurneysen, R. (1892) Die Stellung des Verbums im Altfranzösischen. In *Zeitschrift für Romanische Philologie*, 16, 289-307.
- Vance, B. (1997). *Syntactic Change in Medieval French : Verb-Second and Null Subjects*, Dordrecht, Boston, London : Kluwer Academic Publishers.

¹ L'ancien français présente une autre caractéristique des langues germaniques : l'inversion germanique, dans laquelle, le sujet, nominal ou pronominal, suit immédiatement la forme conjuguée du prédicat verbal, et précède les formes non conjuguées. L'ancien français présente par ailleurs, de manière plus attendue, les caractéristiques d'une langue romane : non expression du sujet et inversion dite « romane ».

² Il n'est pas facile de définir précisément ce qu'est la phrase en français médiéval, et la pertinence même de cette notion peut être remise en question (Lavrentiev, 2010). Dans le présent article, nous utiliserons d'abord ce terme entre guillemets au sens flou de la grammaire traditionnelle avant de donner dans la section 2.1 la définition opératoire adoptée dans le projet d'annotation de notre corpus.

³ Nous utilisons la pagination de l'édition Marchello-Nizia (2011), qui correspond à la numérotation des colonnes dans son manuscrit de base (Lyon, BM, P.A. 77).

⁴ L'auteure déplore, en 2000, l'absence d'un corpus étiqueté suffisamment important, qui permettrait d'étendre son étude à d'autres textes.

⁵ Projet co-financé par l'ANR (France) et la DFG (Allemagne) de 2009 à 2012, coordonné par S. Prévost (Lattice) et A. Stein (Université de Stuttgart).

⁶ Nous utilisons ici les données et les typologies élaborées dans le cadre de la Base de Français Médiéval (<http://bfm.ens-lyon.fr>) et du projet *Corpus représentatif des premiers textes français* (<http://corptef.ens-lyon.fr>) soutenu par l'Agence nationale de la recherche.

⁷ Dans les colonnes du Tableau 1 et des suivants, les textes sont ordonnés par forme (vers / mixte / prose), puis par date de composition.

⁸ Les références bibliographiques des textes du corpus se trouvent en fin d'article.

⁹ Il existe aussi une différence dans le traitement de la coordination entre ce tableau et les suivants. En effet dans le corpus SRCMF, chaque conjoint dans une coordination est traité comme un satellite séparé ; par conséquent, trois sujets coordonnés sont comptés comme trois EPV. Ce tableau sommaire respecte la démarche du corpus. Cependant, dans les tableaux qui suivent, nous avons décidé de nous écarter de cette analyse, en considérant que les coordinations des éléments contigus sans répétition du relateur (préposition ou conjonction de subordination) constituent un seul EPV dans leur ensemble. Cette approche est plutôt associée à une analyse de constituants qu'à une analyse dépendancielle du corpus.

¹⁰ Il faut cependant noter que, en dépit d'une taille inférieure à celle des deux textes en prose, *Beroul* présente un nombre de phrases plus élevé.

¹¹ Ce chiffre reste peu important compte tenu du nombre total de 'Circ Circ' dans *Beroul* (99 occ.), ce qui s'explique par le fait que toutes les combinaisons possibles des trois catégories SUB, ADV, SN se rencontrent dans ce texte.

¹² Il s'agit des modèles suivants 'Circ Circ Sj' (5 occ.), 'Circ Prnt Sj' (3 occ.), 'Prnt Prnt Circ Sj' (2 occ.), 'Prnt Circ Sj' (1 occ.) et 'Cir Prnt Prnt Sj' (1 occ.).

¹³ Il s'agit de 4 occurrences de 'Circ R', 4 occurrences de 'Circ Circ Sj' (cf. tableau 6), et 11 autres structures apparaissant une fois seulement (par ex. 'Circ Prnt Prnt', 'Circ R Circ', etc.).

¹⁴ Le mot utilisé dans les références dans le corps de l'article est indiqué en gras.